

ScholarWorks@GSU

World Language Teacher Candidate Performance on edTPA: An Exploratory Study

Authors	Hildebrandt, Susan A.;Swanson, Peter B.
Download date	2026-03-12 16:03:36
Link to Item	https://hdl.handle.net/20.500.14694/10596

Title: World Language Teacher Candidate Performance on edTPA: An Exploratory Study**Abstract**

Federal and state legislation continues to promote teacher accountability in the United States. The new edTPA, a subject-specific teacher performance assessment, is purported to measure beginning teacher readiness and is being pilot tested and implemented for licensure and certification decisions across the country. In this exploratory quantitative study, the researchers examined edTPA scores of 21 world language teacher candidates from two teacher preparation programs and compared those results to the cut scores for the states of Washington and New York. Results indicated that participants performed best in the planning section and were most challenged by the assessment section. This research has implications for teacher certification candidates, world language teacher preparation programs, policy makers, and other stakeholders.

Key words: pre-service teacher preparation, preparation and certification, program monitoring and assessment, foreign/ second language teacher preparation, student teachers/ interns

Gauging and monitoring teacher effectiveness occupies a critical place in federal and state educational policy. No Child Left Behind (U.S. Department of Education, 2002) brought *highly qualified* into common parlance, and the later Race to the Top required that, in order to receive full federal funding, states needed to measure teacher effectiveness (U.S. Department of Education, 2009). New state-level legislation focuses on teacher preparation and license or certification standards, highlighting teacher performance and effectiveness (Georgia Professional Standards Commission, 2014; Illinois State Board of Education, 2012).

The newly developed edTPA was created by the American Association for Colleges of Teacher Education and the Stanford Center for Assessment, Learning, and Equity (SCALE) to assess new teacher readiness to teach in 27 different content areas (SCALE, 2013). This national standardized assessment, usually carried out during student teaching, was designed to measure teacher candidates' performance to plan, instruct, and assess student learning, focusing on pedagogical skills and informing licensure and certification decisions across the country (SCALE, 2013). However, at the time of this writing, no empirical research exists on the use of

edTPA and its impact on content-specific teacher preparation programs. This exploratory study sought to begin the discussion by examining how teacher candidates from two world language teacher preparation programs scored on edTPA and how those scores compared to known passing scores in two states (New York and Washington) where edTPA has been fully implemented. Specifically, this study sought to use these data to evaluate the two programs and to gain insight into their teacher candidates' ability to succeed on edTPA.¹

Literature Review

Teacher education programs have many stakeholders to whom they are accountable: the U.S. Department of Education, state boards of education, university and college programs and faculty members, accreditation bodies, teacher candidates, future employers and, perhaps most importantly, the teacher candidates' future students. This literature review explores those accountability systems, addresses varying definitions of teacher effectiveness, and compares existing teacher performance assessments, including edTPA.

Accountability for Teacher Education Programs

The last three decades have been characterized by repeated efforts to reform the American educational system. *A Nation at Risk* (National Commission on Excellence in Education, 1983) was perhaps the first in what has become an on-going series of critiques of teacher education. With its focus on content classes and de-emphasis on pedagogical preparation, this report laid the foundation for subsequent discussions of, and proposals for, educational reform (e.g., the Holmes Group, 1986). The *highly qualified teacher* designation was introduced in the *No Child Left Behind* legislation (U.S. Department of Education, 2002), while the Obama administration's *Race to the Top* (U.S. Department of Education, 2009) and *Our Futures, Our*

¹ The authors must point out that they do not have an affiliation with Pearson or SCALE and, therefore, have no self-interest in use of edTPA.

Teachers report (U.S. Department of Education, 2011) reflect the current administration's efforts to reform teacher education policies and practices.

The push for greater teacher accountability directly influences the viability of existing teacher education programs, as current federal teacher education policy seeks to promote programs whose graduates have demonstrated positive impact on student learning and eliminate ineffective programs (Peck, Singer-Gabella, Sloan, & Lin, 2014; U.S. Department of Education, 2011). Thus, individual teacher candidates and the programs that prepare them are pressured to perform or risk failure and ultimately program closure.

State Boards of Education and other educational agencies set teacher certification or licensure standards, evaluate and accredit teacher education programs, and grant individual teacher candidates a teaching license on the recommendation of each teacher preparation institution. State level entities are also making increased efforts to tie student academic performance to individual teachers and to the programs that prepared them (U.S. Department of Education, 2011). To demonstrate their compliance with state and federal policies and to substantiate their success in preparing highly qualified teachers, teacher education programs must design assessments and provide acceptable evidence of teacher effectiveness in annual state reports (Peck & McDonald, 2013). When carefully considered, results from such a range of assessments of teacher candidates' knowledge, skills, and dispositions can be used to enhance the decision-making processes of teacher candidates, faculty members, teacher preparation programs, and extra-programmatic entities (Peck et al., 2014).

Measuring Teacher Effectiveness

Teachers have an undeniable impact on student achievement (Darling-Hammond, 2010a; Ingersoll & Strong, 2011; Stronge, Ward, & Grant, 2011). However, determining exactly what

constitutes effective teaching is notoriously difficult (Muijs, 2006). Assessments of teacher knowledge, rather than assessments of their effectiveness, have often been the norm in state licensure/certification decisions (Darling-Hammond, 2010a). Cochran-Smith, Piazza, and Power (2013) noted, however, that American teacher education as a whole “has made a major programmatic shift from inputs and processes to outcomes” (p. 12). No longer are seat times and completed assignments acceptable evidence for licensure decisions. Rather, teacher candidates must demonstrate “the results of classroom processes, such as impact on student learning” (Goe, Bell, & Little, 2008, p. 4). The Obama administration’s *Race to the Top* further required grantee states to use student learning as evidence in teacher evaluation practices (Darling-Hammond, 2012), although states determine the levels to which teacher candidates must perform. In practical terms, state boards of education or legislatures determine the acceptable cut-scores for the various teacher assessments put into place (Goldhaber, 2007).

To address the inadequacies of more traditional approaches to assessments, performance-based assessments that focus on what teachers actually do in the classroom have begun to complement existing means of measuring teacher readiness. Peck et al. (2014) noted that using standardized performance assessments in teacher education creates “a shared language and a shared agenda for evaluation and improvement of practice” (p. 24). In addition, such shared perspectives and practices can contribute to teacher professionalization by providing a documented foundational knowledge base. Darling-Hammond (2010a) suggested that such assessments of teacher performance can provide specific information about contextualized teacher behaviors and student outcomes. Those assessments can inform stakeholders of the extent to which professional standards are being met, as determined by trained, and possibly more objective, evaluators from beyond a candidate’s home institution. She pointed out that,

in addition to selecting teachers who can indeed teach, these kinds of standards and assessments can help teachers learn to teach more effectively, improve the quality of preparation programs, and create norms that are widely shared across the profession so that good teaching is no longer a magical or haphazard occurrence (2010b, p. 44).

A variety of recent accountability measures have focused on measuring pre-service teacher effectiveness by using K-12 student data derived from in-class assessments. The Teacher Work Sample (Renaissance Partnership for Improving Teacher Quality, 2002) outlined teaching processes and asked teacher candidates to create an assessment plan, provide evidence of instructional decision-making, use student learning to adjust their teaching, interpret data, and communicate with others about students' progress. The new Council on the Accreditation of Educator Preparation/American Council on the Teaching of Foreign Languages standards (Foreign Language Teacher Preparation Standards Writing Team, 2013) for teacher preparation also require teacher candidates to demonstrate their effectiveness. In particular, Standard 5 (Assessment of Language and Cultures - Impact on Student Learning) requires that "candidates reflect on the results of student assessments, adjust instruction accordingly, analyze the results of assessments, and use success and failure to determine the direction of instruction" (Standard 5b, p. 30) and that "candidates interpret and report the results of student performances to all stakeholders and provide opportunity for discussion" (Standard 5c, p. 30). Although Schulz (2000) found that many have suggested that teacher candidates take courses in testing and measurement, today's teacher candidates must use data from their own students to inform and improve both their own teaching and students' learning outcomes.

In addition to changes to initial teacher licensure, state legislation frequently requires that a teacher's development not end at completion of an initial licensure or certification program.

Darling-Hammond (2012) recently advocated for a continuum of authentic teacher performance development opportunities and assessments to monitor and provide support for career-long development. This continuum would, she argued, determine and enhance a teacher's impact on student learning in the long-term. The first such effort to articulate what teachers should know and be able to do began in 1987, with the creation of the National Board for Professional Teaching Standards. The National Board drew on the work of accomplished teachers and educational researchers to determine the standards by which to measure veteran teacher performance and effectiveness (Darling-Hammond, 2012). Such foundational work with assessments for experienced teachers informed the later development of beginning teacher performance assessments like edTPA, the content-specific portfolio assessment.

Implementing edTPA

edTPA is a nationally available performance assessment of beginning teacher readiness, assessing pedagogical knowledge and pedagogical content knowledge. According to Sato's (in press) exploration of the its underlying conception of teaching, edTPA is "fairly neutral on its stance between teacher-centered and student-centered approaches" (p. 7), derived from a somewhat constructivist approach, and aligned with specific disciplinary standards created at local, state, and/or national levels. Evolved from California legislation mandating use of teacher performance assessments (Luster, 2010), edTPA is in various stages of implementation in 34 states and the District of Columbia (American Association for Colleges of Teacher Education, 2014) and is being used to inform initial licensure and certification decisions. Its alignment with ACTFL/CAEP standards is currently being explored by a team put together by ACTFL.

Paralleling the Charlotte Danielson model of teacher evaluation (Sato, in press), edTPA seeks to evaluate the beginning teacher's readiness by assessing three to five lessons created by

the individual teacher candidate within three areas: Planning for Instruction and Assessment, Instructing and Engaging Students in Learning, and Assessing Student Learning. Teacher candidates are assessed through a digital portfolio that includes extensive written passages and videotaped teaching segments. Performance in each of the three areas is scored by trained assessors using standardized rubrics, with each rubric ranging from level 1, the lowest, to level 5, the highest.

Although many states are pilot testing edTPA, determining what teacher assessments to use and the level of acceptable teacher candidate performance on those assessments remains a local phenomenon. Licensure and/or certification criteria, including edTPA cut scores and the acceptable scores across rubrics, are determined at the state level, often by state legislatures (Kornfeld, Grady, Marker, & Ruddell, 2007). Therefore, the scores required on each rubric as well as composite scores vary from state to state. However, at a national standard setting meeting (Stanford Center for Assessment, Learning, and Equity, SCALE, 2014a), a cut score of 42 total points was established for content area assessments with 15 rubrics. This measure was further adjusted to consider a full standard error of measurement lower, thus helping states determine an initial cut score ranging from 37 to 42 total score points, which in some states could be raised as time goes on.

To further complicate the process of setting cut scores for some content areas, including world languages, an adjusted professional performance standard (PPS) must be used to determine a passing score for content areas with a greater or lesser number of rubrics (SCALE, 2014a). Proportional adjustments were used for assessments in content areas with more than, or fewer than, 15 rubrics so as to ensure that the score for each rubric contributes equally to the

total score across all of the academic disciplines and content areas. According to SCALE (2014a),

this PPS was calculated upwards for credential areas with more than 15 rubrics (where a higher total score is possible) and downwards for credential areas with fewer than 15 rubrics (where a lower total score is possible). These calculations in PPSs are proportional to the number of rubrics and maintain the same average rubric score (p. 1).

For example, for world language edTPA, there are 13 rubrics, or two fewer than the suggested number of 15. Each rubric has five levels, and teacher candidates can earn between 1 and 5 points on each rubric. Thus, the maximum score is 65 (i.e., 13 rubrics X 5 performance levels = 65), and suggested cut scores range from 32 to 36 points (SCALE, 2013).

As edTPA was being investigated nationally as a means to measure beginning teacher readiness, numerous criticisms have arisen. First, it is new and little is known about how it compares to existing measures of novice teacher effectiveness and student achievement (Lewis & Young, 2013). Additionally, skeptics cite concerns regarding the involvement of a large corporation, Pearson Inc., to score portfolios. Specifically, Cochran-Smith et al. (2013) stated that Pearson Education's involvement in educational policy and the larger corporatization of the public education sector raise many concerns. Among those, they felt that it contributes to the "deprofessionalization of teacher educators" (p. 17) and decreased local control of teacher preparation and evaluation practices. Madeloni and Gorlewski (2013) argued that edTPA narrows the possibilities of teaching and learning and invites corporate encroachment into education while restricting academic freedom. Further, at a cost of \$300, it adds an additional expense to an already costly teacher certification process, which requires teacher candidates to pay for repeated clinical background checks, state content and pedagogical assessments, and

other fees. Regardless of the criticisms, more than half of the states in the nation have adopted edTPA, warranting empirical study of teacher candidate performance on the new assessment.

Several years ago, both Georgia and Illinois began investigating adopting edTPA as a required part of teacher preparation programs. As of September 1, 2015 both states will require, first, that teacher education programs implement edTPA as an evidence-based assessment of teacher effectiveness (Georgia Professional Standards Commission, 2014; Illinois State Board of Education, 2012) and second, that all teacher candidates pass edTPA in the content area they wish to teach in order to earn state licensure. Cut scores for world language and classical languages have yet to be determined in Georgia (Georgia Professional Standards Commission, 2014) and Illinois (American Association for Colleges of Teacher Education, 2014), much like in other states (SCALE, 2014b). However, the New York State Board of Regents (2013) set the cut score to pass the world language edTPA at a total score of 35 points, which requires an average score of 2.73 on each of the 13 rubrics. The state of Washington set the minimal passing score at 30 points (SCALE, 2014c).

To better inform policy decisions in the states of Illinois and Georgia, this study sought to move beyond a brief exploration of the edTPA for modern and classical languages (Hildebrandt & Hlas, 2013), to investigate world language teacher candidate performance on edTPA, and to answer two research questions:

1. How did this sample of world language teacher candidates score on the 13 rubrics of the edTPA?
2. How do the participants' composite scores compare to the known passing cut scores on the edTPA?

Methods

Institutional contexts

The two authors serve as world language teacher education program coordinators at Illinois State University (ISU) and Georgia State University (GSU), respectively. ISU, located in rural central Illinois, is the oldest public university in the state. It is a moderately large public institution with 19,924 students (approximately 80% Caucasian) enrolled in 43 undergraduate and graduate teacher education programs (ISU, 2014). Illinois State is one of the 10 largest producers of teachers in the United States (American Association of Colleges of Teacher Education, 2014).

GSU, founded in 1913, is a public, urban research institution located in downtown Atlanta. It has a larger overall student population of approximately 32,000 students (38% Caucasian) with more than 1,200 students graduating each year from more than 50 undergraduate and graduate educational programs (GSU, 2014a). Each year more than 500 students graduate as teachers in various content areas (College of Education – GSU, 2014). GSU is the second largest producer of teachers in the state.

While there are obvious differences between these two institutions, there are multiple commonalities. First, ISU and GSU appear to be among the largest world language teacher education programs currently in the United States, with a combined enrollment of approximately 216 students (ISU = 100, GSU = 116). At both institutions, teacher candidates must complete six credits of coursework in pre-K to grade 12 methods of world language instruction, as well as courses in technology integration, reading instruction, general foundations of education, and working with diverse student populations. Coursework and assignments focus on standards-based, proficiency-oriented approaches to instruction and assessment.

At both institutions, teacher candidates are placed for field experiences in a variety of diverse rural, suburban, and urban pre-K to grade 12 schools. Finally, both universities are regionally accredited and earned accreditation from the National Council for Accreditation of Teacher Education, now the Council for the Accreditation of Educator Preparation (CAEP). Because of these very strong core commonalities and since relatively minor academic differences were noted between the two programs, the two samples were combined to form one data set.

Participants

Following Institutional Review Board approval, 21 teacher candidates in the two aforementioned world language teacher education programs agreed to participate in this study in the spring of 2014. The participants represented the total number of spring 2014 student teachers from both institutions and were assessed during their final field placement, typically known as student teaching. The majority of participants were female (86%) and the mean age was 24.88 years (range = 21 to 45 years old). The candidates were predominantly Caucasian (67%) followed by Latinos (22%) and African Americans (11%). Participants were seeking initial certification in French ($n = 1$), German ($n = 1$), or Spanish ($n = 19$). Eighty-one percent reported having studied abroad for an average of four months. Participant demographics were similar to those of in-service teachers nationally in terms of gender, ethnicity, and world language taught (Swanson, 2012). Additionally, parents of the students involved in the teacher candidates' classrooms approved the videotaping of their children and future use of data for research purposes.

Instrument

Aligned with the National Standards for Foreign Language Learning (National Standards for Foreign Language Education Project, 2006) and the Common Core State Standards, the

world language edTPA is described as an authentic assessment tool that shows how teacher candidates develop and evaluate student learning (SCALE, 2013, 2014d). Via a three-step teaching cycle of planning, instruction, and assessment, teacher candidates plan three to five lessons, justify planning decisions, analyze their instruction via video, and use student data to inform their practice. edTPA can be scored locally for formative purposes or can be evaluated externally and officially.

The world language portfolio contains 13 five-point Likert scale rubrics within the three areas of Planning (Rubrics 1, 2, 3 and 4), Instruction (Rubrics 5, 6, 7, 8 and 9), and Assessment (Rubrics 10, 11, 12 and 13). The planning tasks document teacher candidates' *intended* teaching, the instruction tasks document teacher candidates' *enacted* teaching, and the assessment tasks document teacher candidates' *impact* on student learning (SCALE, 2014d). Via Pearson's online system, teacher candidates submit artifacts created by both the teacher candidate and his or her students related to the three areas. Artifacts usually include lesson plans, copies of instructional and assessment materials, video clips of in-class instruction, and student work samples.

As previously mentioned, scores can range from zero to 65 total points in world languages. Each of the 13 rubrics can be scored from 1 to 5 and, according to SCALE's (2013) field tests,

- Level 1 represents the low end of the scoring spectrum, representing the knowledge and skills of a struggling candidate who is not ready to teach;
- Level 2 represents the knowledge and skills of a candidate who is possibly ready to teach;
- Level 3 represents the knowledge and skills of a candidate who is ready to teach;

- Level 4 represents a candidate with a solid foundation of knowledge and skills for a beginning teacher;
- Level 5 represents the advanced skills and abilities of a candidate very well qualified and ready to teach (p. 12).

SCALE has exclusive authorship and copyright for all edTPA handbooks, rubrics, and training/scoring materials, and such specific information cannot be presented here due to copyright restrictions. As mentioned earlier, individual states set their own passing scores for the various content areas and have the authority to alter cut scores over time.

Procedures and Data Analysis

Trained Pearson evaluators scored all ISU portfolios via a LiveText interface with Pearson, and the ISU Office of the Provost paid each portfolio's \$300 fee. At GSU, teacher preparation program coordinators were encouraged to pilot test and locally assess candidates' portfolios following SCALE training. All GSU teacher candidate portfolios were turned in via LiveText and locally scored by the second author and a GSU colleague, who were trained by SCALE to evaluate edTPA assessments. A high level of inter-rater reliability was found as the GSU evaluators only disagreed on three of the 65 total rubrics' ratings. Also, it is also important to note that no instructor or supervisor support was provided to candidates as they worked on their official edTPA portfolio beyond time scheduled during an on-campus meeting for teacher candidates to critique each other's submissions.

The numerical ratings for participants' scores on the 13 edTPA rubrics and their demographic data were entered into and analyzed using SPSS 19.0 during April and May 2014. Due to the low number of participants, only frequency counts, means, and standard deviations are reported here.

Findings

Teacher candidates' scores on each of the 13 world language rubrics on the edTPA are reported in Table 1.

Table 1

<i>Means, standard deviations, and frequencies of each performance level on the 13 edTPA rubrics.</i>							
	<i>M</i>	<i>SD</i>	1	2	3	4	5
Planning							
- Planning for Communicative Proficiency in the Target Language (R1)	3.86	.65	0	1	3	15	2
- Planning to Support Varied Student Learning Needs (R2)	3.57	.59	0	1	7	13	0
- Using Knowledge of Students to Inform Teaching and Learning (R3)	3.52	.75	0	1	10	8	2
- Planning Assessments to Monitor and Support Students' Development of Communicative Proficiency in the Target Language (R4)	3.62	.67	0	1	7	12	1
Instruction							
- Learning Environment (R5)	3.71	.72	0	0	9	9	3
- Engaging Students' Target Language Communication (R6)	3.43	.67	0	2	8	11	0
- Deepening Student Communicative Proficiency in the Target Language (R7)	3.24	.83	0	4	9	7	1
- Subject-Specific Pedagogy (R8)	3.05	.92	0	4	10	5	1
- Analyzing Teaching Effectiveness (R9)	3.24	.70	0	3	10	8	0
Assessment							
- Analysis of Student Communicative Proficiency in the Target Language (R10)	3.38	.80	0	4	5	12	0
- Providing Feedback to Guide Student Development of Communicative Proficiency in the Target Language (R11)	3.19	1.07	2	4	3	12	0
- Student Use of Feedback (R12)	2.71	1.05	2	8	6	4	1
- Using Assessment to Inform Instruction (R13)	2.90	.62	1	2	16	2	0

These data provide insight into three core aspects of professional pedagogical content knowledge. Data showed a range of scores for these teacher candidates across the four rubrics addressing the first core area, Planning. Specifically, the participants scored the highest on Rubric 1: Planning for Communicative Proficiency in the Target Language. Data indicated that, in their lesson plans, the participants were able to make connections between language forms and functions within a meaningful cultural context ($M = 3.86$, $SD 0.65$), but their planning lacked focus on all three modes of communication (SCALE, 2013b). A slightly lower mean was found on Rubric 4: Planning Assessments to Monitor and Support Students' Development of Communicative Proficiency in the Target Language ($M = 3.62$, $SD 0.67$). However, on both of the aforementioned rubrics the majority of the participants were still rated at the third and fourth highest performance levels. The two lowest mean scores were found on Rubric 2: Planning for Varied Student Learning Needs and Rubric 3: Using Knowledge of Students to Inform Teaching and Learning, although only one participant scored in the two lowest performance levels on these two rubrics.

Table 1 also reflects findings for rubrics 5, 6, 7, 8 and 9 that addressed the second core aspect of pedagogical content knowledge, Instruction. Participants scored the highest ($M = 3.71$, $SD 0.72$) on Rubric 5: Learning Environment. This indicates that these teacher candidates provided a low risk, social environment that challenged students to express themselves. On Rubric 5, none of the 21 participants scored in the two lowest performance levels. The second highest mean score was found on the Rubric 6: Engaging Students' Target Language Communication, for which the majority of the participants scored in the third and fourth performance levels (93%) and for which none of the participants' performances was rated at the lowest or the highest ends of the rubric. This finding indicates that the participants demonstrated

an ability to engage learners in linking prior learning to new learning, but that they did not lead language learners to deepen and extend communicative proficiency in the target language (SCALE, 2013b).

Slightly lower means were found for Rubrics 7, 8 and 9, for which approximately one-fifth (17%) of the participants' performances were rated in the two lowest performance levels on each of the three rubrics. Examination of participant performance on Rubric 7: Deepening Student Communicative Proficiency in the Target Language showed that 33% of the participants were able to prompt and build on students' responses in order to develop communicative proficiency (SCALE, 2013b). The lowest mean score was found on Rubric 8: Subject-specific Pedagogy ($M = 3.05$, $SD = 0.92$), indicating that the teacher candidates in this study had some difficulties providing opportunities for students to make comparisons and connections between their prior experiences and knowledge and the new cultural practices, products, and perspectives (SCALE, 2013b). A slightly higher mean was found on Rubric 9: Analyzing Teaching Effectiveness ($M = 3.24$, $SD = 0.70$), indicating that these teacher candidates had some difficulty using evidence to evaluate and modify their instructional strategies to meet their students' learning needs (SCALE, 2013b).

Examination of teacher candidate performance in Assessment, using Rubrics 10, 11, 12 and 13, revealed that participants scored the highest on Rubric 10: Analysis of Student Communicative Proficiency ($M = 3.38$, $SD = 0.80$) with more than half (57%) of the participants scoring in the second highest performance level on the rubric. This finding indicated that they were able to identify patterns in student learning when analyzing student data. Similar results were found for student performance on Rubric 11: Providing Feedback ($M = 3.19$, $SD = 1.07$). Participants' lowest mean scores for this core domain were found on Rubric 12: Student Use of

Feedback ($M = 2.71$, $SD = 1.05$), which was the lowest mean across all of the 13 rubrics, and Rubric 13: Using Assessment to Inform Instruction ($M = 2.90$, $SD = 0.62$). Closer examination of the frequencies for each performance level for these two rubrics showed that approximately one-third of the participants scored in the two lowest performance levels on Rubric 12, and 76% percent of the participants were rated in the middle of the 5-point rubric (performance level 3) on Rubric 13. Furthermore, on Rubric 13, only two of the participants were able to provide targeted support to learners in order to improve their communicative proficiency related to the interpretive mode and at least one of the other two modes of communication, as demonstrated by their score of 4 for the rubric (SCALE, 2013b).

Overall, candidates were most successful in the Planning tasks and least successful in the Assessment tasks, with participants scoring highest in the area of Planning ($M = 3.64$, $SD = 0.46$). The teacher candidates in this study performed slightly lower on average on the five rubrics constituting the Instruction subgroup ($M = 3.33$, $SD = 0.56$), with Assessment as the lowest of the three areas ($M = 3.04$, $SD = 0.96$). From a collective perspective, teacher candidates demonstrated skills that approached level 3 of the 5-point rubric, which “represents the knowledge and skills of a candidate who is ready to teach” (SCALE, 2013, p. 1).

The second research question investigated how the participants’ composite scores compared to the known passing cut scores on edTPA. As shown in Table 2, participants’ composite scores were compared to the cut scores for the states of Washington and New York.

Table 2

Means and standard deviations for the three areas and the total scores of the edTPA.

	<i>M</i>	<i>SD</i>
Planning	3.64	.46
Instruction	3.33	.56
Assessment	3.04	.96

Total edTPA score	43.12	5.98
-------------------	-------	------

For this sample of teacher candidates, composite scores on the edTPA ranged from 31 to 52 total points with a mean of 43.12 ($SD = 5.98$) out of a possible 65 points. Two participants scored in the 30 to 35 point range, five in the 35 to 40 point range, four in the 40 to 45 point range, seven in the 45 to 50 point range, and three above 50 points. Given these results, all of the participants would have passed the edTPA in the state of Washington, which has a cut-score of 30, and 90% ($n = 19$) would have met or exceeded the cut score of 35 in New York.

Discussion

This pilot project sought to better understand how a sample of world language teacher candidates in Georgia and Illinois scored on the new national assessment, edTPA in world languages, and how their scores compared to those cut scores already set in two other states, New York and Washington. As the first empirical exploration of edTPA in any content area, this study sought to begin a healthy conversation about the new student teacher assessment and its potential impacts at the local, state, and national levels. The following discussion will explore possible reasons for participants' success and suggestions for other programs implementing edTPA. It will also examine edTPA's place within CAEP's paradigm of world language teacher preparation program accreditation, the establishment of states' cut scores, edTPA's impact on the quality and quantity of beginning world language teachers, and the rising cost of world language teacher education.

Data showed that the 21 participants in this study scored the highest on the Planning tasks and were most challenged by the Assessment tasks. This finding was perhaps due to the fact that world language teacher candidates, and possibly all teacher candidates, may have the most experience planning for lessons during their education coursework and their content-specific

methods classes. Also, in many certification programs, teacher candidates have opportunities to teach some of those lessons to peers or students at their practicum sites, gaining practical experiences that they can later draw on as they carry out the edTPA assessment. Teacher candidates in these two teacher education programs practiced their instructional skills via micro teaches taught to both peers and language learners during in-class field experiences, and feedback from the instructor and peers was used to improve teaching skills. Because language learning can be stressful (Krashen, 1981), teacher candidates were continually reminded not only to “recognize the presence of foreign language anxiety in language learners but also help learners acknowledge, cope with, and reduce their anxiety” (Huang & Eslami, 2010, p. 32). Early in teacher preparation, these teacher candidates learned that by creating a low anxiety learning environment, their students would be more likely to engage in risk-taking behavior with regard to practicing and using a second language (Krashen, 1981, 1985).

Where they tended to demonstrate a lack experience, however, is in the area of using student performance data to inform their own teaching. Rarely, if ever, are teacher candidates in a position that allows them to have sustained contact with students whose work they could use to inform future teaching of the same students. While the curricula for both programs in this study focus heavily on planning and instruction and purposely have two methods classes, neither program has a required class that specializes in, or places a strong emphasis on, assessment. Assessment has been considered a strand interwoven throughout each program, with explicit instruction limited to readings from textbooks (Sandrock, 2010; Shrum & Glisan, 2010) and in-class discussions. Although thorough preparation in assessment and evaluation have been historically suggested (Schultz, 2000), the participants’ teacher preparation programs do not consistently offer a stand-alone assessment class. What is more, adding a required assessment

class to already tight schedules and programs of study could prove very challenging, particularly since teacher education programs in Georgia are limited to 120 total credit hours (University System of Georgia, 2014). Thus, adding credits to existing programs can extend time to graduation and employment.

Establishing timelines with explicit and carefully thought through deadlines for completing the edTPA portfolio may also support candidates' success on edTPA. It is important that teacher candidates have adequate time to complete the portfolio, submit it for evaluation, and still have time to revise and resubmit sections that earned an unsatisfactory score, if necessary and at an additional cost to the candidate. At present, the timeline during which most teacher candidates complete the edTPA is less than one semester. That is, teacher candidates must submit the portfolio for external evaluation slightly past the midpoint of the semester and then wait approximately three weeks for the results. If any part is deemed unsatisfactory by the external evaluators, the teacher candidate has a limited amount of time to submit a different artifact for evaluation with the hope of receiving a satisfactory score. Clearly, at these two universities, the current timeline for edTPA evaluation may be problematic. Requiring candidates to submit their work early enough in the student teaching experience to allow for evaluation and subsequent resubmission, if needed, may threaten candidates' performance because they may not have gained sufficient mastery during the first half of their student teaching experience to succeed on this high stakes assessment.

In addition, world language teacher preparation programs, such as the two under consideration here, may have other institutional or state rules that govern the length and/or the beginning and ending dates of candidates' student teaching experiences. To begin to resolve scheduling and submission issues, both program coordinators in this study are questioning the

requirement that student teachers wait to begin their practicum experience based on universities' calendars, rather than beginning student teaching on the first day of the K-12 calendar during the fall and spring semesters. In Georgia, for example, local school districts begin the first week of August and GSU classes begin three weeks later. By following the school district calendar as a starting point, teacher candidates may be provided with sufficient additional time during which to gain much-needed experience in the classroom, as well as to prepare their portfolios prior to turning them in about the ninth week of their student teaching assignments. At present, SCALE estimates that evaluation will take approximately three weeks. Therefore, by adding additional weeks to the beginning of the experiences and by having teacher candidates complete and turn in the portfolios for evaluation slightly beyond the midpoint of the semester, time remains so teacher candidates can revise any portions that may require additional attention and subsequent reevaluation.

In addition to establishing timelines and extending the student teaching experience, program coordinators can further support candidates' success by helping candidates to become more familiar with the assessments themselves. For example, teacher candidates from ISU completed an abbreviated edTPA, based on teaching K-4th graders at a local community center, so that candidates could gain familiarity with the assessment and acquire authentic student data to analyze prior to the high stakes submission of an officially scored edTPA portfolio. Program directors may want to consider implementing such early preparation with the edTPA in field practica coursework. In addition, with consent from previous teacher candidates, written edTPA responses and videotaped lessons from former teacher candidates may be examined by current cohorts of teacher candidates in methods and practicum classes so to gain insight into the assessment and the skills necessary to be successful. Care must be taken, however, to use these

samples as learning opportunities instead of limiting teacher preparation to preparing for one assessment or teaching to the test.

It is also important to consider how candidates' portfolios can be used as part of the program's accreditation report. With the release of the new ACTFL /CAEP accreditation standards in 2013, work is underway to establish crosswalks between those standards and edTPA. The authors speculate that various elements of the edTPA portfolio will be permitted to serve as one or more of the six to eight key assessments required for program accreditation, providing acceptable evidence of teacher candidates meeting Standard 3 (Language Acquisition Theories and Knowledge of Students and Their Needs), Standard 4 (Integration of Standards in Planning, Classroom Practice, and Use of Instructional Resources), and Standard 5 (Assessment of Languages and Cultures – Impact on Student Learning). Using parts from the edTPA in order to document teacher candidates' skills would complement already existing evidence, although program directors are advised to carefully audit required assessments and eliminate those that prove repetitive or uninformative.

However, while the portfolios can be used for several other purposes, it must be remembered that their primary purpose is to inform licensure or certification decisions. This study's participants performed successfully on edTPA, with composite scores that would support certification or licensure in the only two states with determined cut scores for world language teacher candidates. While some in the profession express concerns about edTPA (e.g., Mandeloni & Gorlewski, 2013), the findings of this study do not provide evidence its expectations of student teachers are excessively high. Results from this study should be encouraging to teacher candidates and program coordinators as edTPA becomes consequential in their states. While Georgia and Illinois have not yet set cut scores for the world language edTPA,

all of the teacher candidates would have passed if they sought certification in Washington state, and 19 of the 21 candidates would have passed if they sought certification in New York. Given the local nature of the American educational system and teacher licensure or certification decisions (Kornfeld, Grady, Marker, & Ruddell, 2007), each state that chooses to adopt edTPA will ultimately determine cut scores as part of their teacher licensure or certification process. Some states, like Illinois, have established cut scores that gradually ascend over time so as to allow teacher preparation programs to gradually prepared each successive group of teacher candidates to meet the more exigent requirements. Cut scores have not yet been announced for content areas like world languages that do not use the common 15 rubric format. At present, only New York and Washington state have established pass scores for edTPA, and neither state has released the way in which those scores were determined. Other states, like Georgia, are in the process of pilot testing edTPA, but because edTPA is so new, there is no published research at the time showing pass rates or scores.

With the dearth of empirical data and edTPA's impact on educational systems, policy makers should carefully determine edTPA cut scores in order to balance quality and quantity of beginning teachers, especially in states like Georgia that are currently experiencing a shortage of world language teachers. For years, there has been a shortage of world language teachers (Swanson, 2013), and the implementation of edTPA could aggravate this shortage. For example, if passing cut scores on edTPA are set too high, fewer teacher candidates may become certified, exacerbating the world language teacher shortage, particularly since world language teacher education programs already tend to have low enrollments (personal communication, Judith Shrum, September 10, 2014). In contrast, if the cut score is set too low, the teacher shortage may ease a little at the expense of certifying less qualified language teachers. While states struggle to

find enough certified language teachers, research shows that hiring non-qualified instructors, as well as instructors who enter the profession through alternate routes of certification, results both in less effective teachers than those who pass through traditional routes as well as in higher professional attrition rates (Darling-Hammond, 2010c; Darling-Hammond, Berry, & Thoreson, 2001). Thus, careful attention must be taken when establishing cut scores.

In addition to initially setting edTPA cut scores, policymakers need to take into account the burden of rising costs of becoming certified to teach. As college education becomes increasingly expensive (McPherson, 2010), student teachers already accumulate tuition debt, which most certainly increases during the unpaid student teaching experience during which they are generally strongly warned against working at other, paying positions. Some question the frequency of high stakes tests of basic skills tests, state tests of content and pedagogical knowledge, professional ethics, and mandated teacher performance assessments, like edTPA. For example, in Georgia, teacher candidates must have at least one background check (\$49.50) and tort liability insurance (\$7) in order to be eligible for field placements in schools (Georgia State University, 2014b). Additionally, they must pay the following amounts in order to receive a teaching certificate: \$128 for the GACE Program Admission Assessment (basic skills), \$193 for the Content Pedagogy Assessment, and \$60 for the Georgia Educator Ethics Assessment (Georgia Professional Standards Commission, 2014). Additionally, teacher candidates must have a LiveText account (\$80). Then, once in student teaching, they must pay \$300 for edTPA. Should a teacher candidate not receive a passing score on one task, the individual must redo that task and pay an additional \$100 evaluation fee. If the entire edTPA must be retaken, the cost is another \$300. Furthermore, once the teacher candidate has passed the examinations, the individual must then apply for certification and pay an additional \$20 in Georgia. In Georgia's

case, a tiered certification system is being put into place, such that, if teacher candidates do not pass the state tests at the Professional level, they receive an Induction certificate that expires in three years, after which they must pay for, and take, the exams again in order to apply for a Professional certificate. Thus, the current total cost of the required assessments, assuming that candidates pass each of them on the first attempt, is \$817.50. Finally, unlike Georgia, other states, like Wisconsin, for example, also require teacher candidates to pass the Oral Proficiency Interview (OPI), at a cost of \$139 (Language Testing International, 2014), as do all candidates in teacher preparation programs seeking CAEP accreditation. Clearly, adding additional formal assessment experiences like edTPA to existing assessment mandates may further discourage prospective teachers, who already struggle with sizable educational debt while preparing for a job that is compensated at about \$36,000 annually at the beginning of a career.

While this exploratory study shed light on edTPA outcomes, it is not without its limitations. While the number of participants is low, it must be acknowledged that world language teacher education programs are generally small (personal communication, Judith Shrum, September 10, 2014). Therefore, a sample of 21 participants, while modest, accounts for all teacher candidates who student taught during the spring of 2014 in both universities. Additionally, edTPA portfolios from GSU participants were scored by evaluators from their home institution and were not officially or externally reviewed. Despite all efforts to contain bias, had trained evaluators from Pearson officially scored those portfolios, scores may have differed.

Conclusions

From the moment that No Child Left Behind was enacted in 2001 and Race to the Top bolstered that initial mandate, measuring teacher effectiveness became a high educational

priority. Results from the present study suggest that many teacher candidates are, in fact, being prepared to meet the challenges posed by legislation that mandates K-12 teacher accountability. edTPA has been already adopted, although is not yet implemented, in a number of states, and it is in the best interest of teacher preparation programs across the disciplines to consider the impact of its implementation. An exploration of cut scores and passing rates, how teacher candidates and cooperating teachers perceive edTPA, its potential impact on student teaching placements, and best practices for preparing candidates to be successful would be informative. Finally, it would be helpful to know more about how programs provide remediation to teacher candidates who are initially unable to pass one or more sections of the edTPA.

Acknowledgement

This research was supported by a 2013 ACTFL Research Priorities Projects – Phase II grant and explored Research Priority Area #1: FL Teacher Preparation Model Programs: Documentation, Implementation, and Outcomes.

References

American Association for Colleges of Teacher Education. (2014). *Participation map*. Available from, <http://edtpa.aacte.org/state-policy>

Cochran-Smith, M., Piazza, P., & Power, C. (2013). The politics of accountability: Assessing teacher education in the United States. *The Educational Forum*, 77, 6-27.

College of Education – GSU (2014). *About the college of education*. Available from, <http://education.gsu.edu/about-the-college-of-education/>

Darling-Hammond, L. (2010a). Teacher education and the American future. *Journal of Teacher Education*, 61, 35-47.

Darling-Hammond, L. (2010b). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*. Washington, DC: Center for American Progress. Available from, <http://www.americanprogress.org/issues/education/report/2010/10/19/8502/evaluating-teacher-effectiveness/>

Darling-Hammond, L. (2010c). Recruiting and retaining teachers: Turning around the race to the bottom in high-need schools. *Journal of Curriculum and Instruction*, 4(1), 16-32.

Darling-Hammond, L. (2012). The right start: Creating a strong foundation for the teaching career. *Phi Delta Kappan*, 94(3), 8-13.

Darling-Hammond, L., Berry B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, 23(1), 57-77.

Foreign Language Teacher Preparation Standards Writing Team. (2013). *ACTFL/CAEP program standards for the preparation of foreign language teachers*. Available from, <http://www.actfl.org/sites/default/files/pdfs/ACTFL-Standards20Aug2013.pdf>

Georgia Professional Standards Commission. (2014). *Understanding the 2014 educator certification rule*. Available from, http://www.gapsc.com/Commission/policies_guidelines/Downloads/2014EducatorCertificationRuleChanges.pdf

Georgia State University. (2014a). *Quick facts*. Available from, <http://www.gsu.edu/about/quick-facts/>

Georgia State University. (2014b). Applying to the program & field placements. Available from, <http://mcl.gsu.edu/home/undergraduate/bachelor-of-arts/teacher-certification/applying-program-field-placements/>

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, 42(4), 765-794.

Hildebrandt, S. A., & Hlas, A. (2013). Assessing the Effectiveness of New Teachers with edTPA™. *Language Educator*, 8(3), 50-51. Available from http://www.actfl.org/sites/default/files/pdfs/TLE_pdf/TLE_Apr13_TOC.pdf

Holmes Group. (1986). *Tomorrow's teachers*. East Lansing, MI: Author.

Huang, S., & Eslami, Z. (2010). The relationship between teacher and peer support and English language learners anxiety. *English Language Teaching*, 3, 32–40.

Illinois State Board of Education. (2012). Memo. Available from, <http://www.isbe.state.il.us/licensure/pdf/higher-ed/edTPA/edTPA-summary1212.pdf>

Illinois State University. (2014). *School of Teaching and Learning. Welcome from the director*. Available from, <http://education.illinoisstate.edu/about/departments/tch/>

Ingersoll, R., & Strong, M. (2011). The impact of mentoring and induction programs for beginning teachers: A critical review of the research. *Review of Education Research*, 81(2), 201-233.

Kornfeld, J., Grady, K., Marker, P., & Ruddell, M. (2007). Caught in the current: A self-study of state-mandated compliance in a teacher education program. *The Teachers College Record*, 109(8), 1902-1930.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.

Language Testing International. (2014). Oral proficiency interview. Available from, <http://www.languagetesting.com/oral-proficiency-interview-opi>

Lewis, W. D., & Young, T. V. (2013). The politics of accountability: Teacher education policy. *Educational Policy*, 27(2), 190-216.

Luster, J. (2010). Why states should require a teaching performance assessment and a subject matter assessment for a preliminary teaching credential. *Research in Higher Education*, 8, 1-16.

Madeloni, B., & Gorlewski, J. (2013). Wrong answer to the wrong question: Why we need critical teacher education, not standardization. *Rethinking Schools*, 27(4). Available from, http://www.rethinkingschools.org/archive/27_04/27_04_madeloni-gorlewski.shtml

McPherson, M. S. (2010). *Keeping college affordable: Government and educational opportunity*. Brookings Institution Press.

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 12(1), 53-74.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the Nation and the Secretary of Education, United States Department of Education*. Washington, D.C.: National Commission on Excellence in Education.

National Standards in Foreign Language Education Project. (2006). *National standards for foreign language learning in the 21st century* (2nd ed.). Lawrence, KS: Allen Press Inc.

New York State Board of Regents. (2013). *Commissioner King announces scores needed to pass new edTPA exam*. Available from, <http://www.highered.nysed.gov/edtpa.html>

Peck, C. A., & McDonald, M. (2013). Creating “cultures of evidence” in teacher education: Context, policy, and practice in three high-data-use programs. *The New Educator*, 9(1), 12-28.

Peck, C. A., Singer-Gabella, M., Sloan, T., & Lin, S. (2014). Driving blind: Why we need standardized performance in teacher education. *Journal of Curriculum and Instruction*, 8(1), 8-30.

Renaissance Partnership for Improving Teacher Quality. (2002). *Teacher work sample: Performance prompt, teaching process standards, scoring rubrics*. Bowling Green, KY: Author. Available from, <http://www.wku.edu/rtwsc/documents/prompt-and-rubric.pdf>

Sandrock, P. (2010). *The keys to assessing language performance: A teacher's manual for measuring student progress*. Alexandria, VA: ACTFL.

Sato, M. (in press). What is the underlying conception of teacher of the edTPA? *Journal of Teacher Education*.

Schulz, R. A. (2000). Foreign language teacher development: *MLJ* perspectives – 1916-1999. *The Modern Language Journal*, 84, 495-522.

Shrum, J. L., & Glisan, E. W. (2010). *Teacher's handbook: Contextualized language instruction* (4th ed.). Boston, MA: Heinle Cengage Learning.

Stanford Center for Assessment, Learning, and Equity. (2013). *2013 edTPA field test: Summary report*. Stanford, CA: Stanford University. Available from, <http://edtpa.aacte.org/news-area/announcements/edtpa-summary-report-is-now-available.html>

- Stanford Center for Assessment, Learning, and Equity. (2014a). *Recommended professional performance standard*. Available from, http://www.edtpa.com/PageView.aspx?f=GEN_PerformanceStandard.html
- Stanford Center for Assessment, Learning, and Equity. (2014b). *Participation map*. Available from, <http://edtpa.aacte.org/state-policy>
- Stanford Center for Assessment, Learning, and Equity. (2014c). *edTPA for Washington*. Available from, http://www.edtpa.com/PageView.aspx?f=GEN_Washington.html
- Stanford Center for Assessment, Learning, and Equity. (2014d). *World languages handbook*. Stanford, CA: Stanford University.
- Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, 62(4), 339-355.
- Swanson, P. (2012). Second/foreign language teacher efficacy: Multiple factors and their relation to professional attrition. *Canadian Modern Language Review*, 68(1), 78-101.
- Swanson, P. (2013). *Identifying and recruiting language teachers: A research-based approach*. Charlotte, NC: Information Age Publishing.
- University System of Georgia. (2014). *Academic programs: 2.3.5 degree requirements*. Available from, http://www.usg.edu/academic_affairs_handbook/section2/C731
- U.S. Department of Education. (2002). No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425. Available from, <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- U.S. Department of Education. (2009). *Race to the Top executive summary*. Available from, <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- U.S. Department of Education. (2011). *Our future, our teachers: The Obama administration's plan for teacher education reform and improvement*. Available from, <http://www.ed.gov/sites/default/files/our-future-our-teachers.pdf>